

AD-A192 100



DTIC FILE COPY

**COWLES FOUNDATION FOR RESEARCH IN ECONOMICS
AT YALE UNIVERSITY**

Box 2125, Yale Station
New Haven, Connecticut 06520

**DTIC
ELECTE
MAR 02 1988**
S D

COWLES FOUNDATION DISCUSSION PAPER NO. 861

Note: Cowles Foundation Discussion Papers are preliminary materials circulated to stimulate discussion and critical comment. Requests for single copies of a Paper will be filled by the Cowles Foundation within the limits of the supply. References in publications to Discussion Papers (other than acknowledgment that a writer had access to such unpublished material) should be cleared with the author to protect the tentative character of these papers.

**A CENTERED PROJECTIVE ALGORITHM
FOR LINEAR PROGRAMMING**

by

Michael J. Todd
Cornell University

and

Yinyu Ye
Stanford University

February 1988

DISTRIBUTION STATEMENT A

Approved for public release;
Distribution Unlimited

88 2 25 107

A CENTERED PROJECTIVE ALGORITHM
FOR LINEAR PROGRAMMING

by

Michael J. Todd*

and

Yinyu Ye[†]



Accession For	
NTIS CRA&I	<input checked="" type="checkbox"/>
DTIC TAB	<input type="checkbox"/>
Unannounced	<input type="checkbox"/>
Justification	
By <i>per HP</i>	
Distribution /	
Availability Codes	
Dist	Avail. and/or Special
<i>A-1</i>	

*School of Operations Research and Industrial Engineering, Cornell University. Part of this work was completed while visiting the Cowles Foundation for Research in Economics, Yale university, whose support is gratefully acknowledged. Research also supported in part by NSF Grant ECS-8602534 and ONR Contract N00014-87-K-0212.

[†]Department of Engineering-Economic Systems, Stanford University and Integrated Systems Inc., CA. Research completed while visiting Cornell University, with support from the U. S. Army Research Office through the Mathematical Science Institute of Cornell University.

ABSTRACT

(the square root of $n \times L$)

We describe a projective algorithm for linear programming that shares features with Karmarkar's projective algorithm and its variants and with the path-following methods of Gonzaga, Kojima-Mizuno-Yoshise, Monteiro-Adler, Renegar, Vaidya and Ye. It operates in a primal-dual setting, stays close to the central trajectories, and converges in $O(\sqrt{n} L)$ iterations like the latter methods. (Here n is the number of variables and L the input size of the problem.) However, it is motivated by seeking reductions in a suitable potential function as in projective algorithms, and the approximate centering is an automatic byproduct of our choice of potential function.

//

1. INTRODUCTION

This paper is concerned with interior algorithms for the linear programming problem

$$\begin{array}{ll} \min g^T x & \\ (P_1) & Ax = b \\ & x \geq 0, \end{array}$$

where A is a real $k \times n$ matrix, b is in R^k , and g and x are in R^n . Apart from methods related to the ellipsoid algorithm, we can distinguish two classes of such methods:

(i) projective algorithms: Karmarkar's method [14] and its variants including the affine-scaling variant; and

(ii) path-following algorithms, due to Gonzaga [12], Kojima, Mizuno and Yoshise [15,16], Monteiro and Adler [21,22], Renegar [23], Vaidya [28] and Ye [30].

We describe here a method that shares features with both approaches.

Assume that the data A , b and g of (P_1) are integer, and the input size is L . Then Karmarkar's algorithm can obtain within $O(nL)$ iterations a sufficiently accurate solution that an exact solution can easily be deduced. The work required in each iteration is $O(n^3)$ arithmetic operations in the basic version, or an average of $O(n^{2.5})$ operations in the modified algorithm, giving a complexity of $O(n^{3.5}L)$. Each iteration seeks a constant reduction in Karmarkar's potential function; after making a projective transformation, a step is taken in a direction that is the negative of the projected gradient of the potential function in the transformed space. This constant reduction implies the bound of $O(nL)$ for the number of iterations. In practice, far fewer iterations are required; it seems that $O(L)$, or perhaps $O((\ell n n)L)$, suffice when a reasonable line

search is performed. On the other hand, it is known that a reduction of the potential function greater than some fixed constant cannot always be achieved; see Anstreicher [2] and McDiarmid [17].

Considering the infinitesimal step version of Karmarkar's algorithm and its variants, one is led to the study of paths or trajectories in the interior of the feasible region. These paths have been investigated by Bayer and Lagarias [5], Megiddo [18] and Megiddo and Shub [19]. They are closely related to the paths associated with the classical barrier function method of Frisch [11] and Fiacco and McCormick [10], and also to the notion of "analytic center" of Sonnevend [24]. Renegar [23] gave an algorithm using Newton's method to trace the path that required $O(\sqrt{n} L)$ iterations, each needing $O(n^3)$ arithmetic operations. Using Karmarkar's trick of solving approximate systems of equations and making rank 1 updates, Vaidya [28] showed that an average of $O(n^{2.5})$ arithmetical operations per iteration sufficed, giving an overall complexity of $O(n^3 L)$. Gonzaga [12] independently and simultaneously obtained the same result. All these methods operated in the primal space alone.

Kojima, Mizuno and Yoshise [15] described a primal–dual interior algorithm that followed central trajectories in both the primal and dual feasible regions. This method used $O(nL)$ iterations. Monteiro and Adler [21] gave a primal–dual method that required only $O(\sqrt{n}L)$ iterations, and had a complexity of $O(n^3 L)$. A similar complexity was established by Kojima et al. [16] for a path–following method for certain linear complementarity problems, including those arising from linear and convex quadratic programming. Monteiro and Adler also extended their method to one for convex quadratic programming with complexity $O(n^3 L)$ [22]. All these methods work in a primal–dual setting, although there is no additional computation compared to a primal–only method. Finally, Ye [30] has given a primal–only convex quadratic programming algorithm that requires only $O(\sqrt{n} L)$ iterations.

We aim to shed light on these two classes of algorithms by introducing and analyzing a centered projective algorithm for linear programming. Like Karmarkar's method, it is motivated by seeking a reduction in a suitable potential function, and the search direction is the negative of the projected gradient of such a function in a transformed space. Like the path-following methods, it requires only $O(\sqrt{n} L)$ iterations, operates in a primal-dual setting, and stays close (automatically) to the central trajectories. We analyze our method with exact projections, so that it needs $O(n^3)$ arithmetic operations at each iteration. However, we shall see that the direction generated coincides with those of Monteiro-Adler and Kojima et al., so their analysis shows that inexact projections could be employed to give an overall complexity of $O(n^3 L)$. We note that Ye and Todd [31] have shown that the path-following algorithms maintain a reasonable decrease in a certain potential function; this function differs slightly from the one chosen in this paper.

One question we particularly wish to illuminate is the following. The projective methods define a search direction, and the analysis then shows that a constant reduction in the potential function can be achieved; as we have indicated, typical behavior in practice is much more encouraging. On the other hand, the path-following methods try for an *a priori* determined reduction of $(1 - \gamma/\sqrt{n})$ in the duality gap for some constant γ and that is what they obtain. Can one hope to do better, and where does the \sqrt{n} term come from? Some suggestions have been given in earlier papers; we hope our approach yields additional insight.

The paper is organized as follows. In section 2 we describe Karmarkar's algorithm briefly and indicate why a constant reduction in potential is obtained. We also outline the step determination procedure of a path-following method. Section 3 reformulates the primal and dual problems in a symmetric way and introduces a combined primal-dual problem.

In section 4 we state the potential functions with which we will be concerned and discuss scalings of the primal and dual variables. Section 5 computes the necessary projection matrix and hence the search direction used. We also relate this search direction to the step employed in path-following methods. As is perhaps not too surprising in view of earlier analyses by Gonzaga [13] and Mitchell and Todd [20], our direction is a linear combination of the same two directions that arise in path-following algorithms. However, we stress that the motivation behind it is completely different.

Section 6 describes the algorithm which is analyzed in section 7. Finally section 8 contains further discussion.

2. AN OUTLINE OF PROJECTIVE AND PATH-FOLLOWING METHODS

Our problem is

$$\begin{array}{ll}
 \min g^T x & \\
 (P_1) & Ax = b \\
 & x \geq 0,
 \end{array}$$

where A is $k \times n$, b a k -vector, and g and x lie in R^n . We assume

- (A1) (P_1) has a strictly positive feasible solution;
- (A2) The set of optimal solutions of (P_1) is nonempty and bounded; and
- (A3) The matrix A has full row rank k .

To describe Karmarkar's projective algorithm, we assume for simplicity that no optimal solutions lie in the relative interior of the feasible region of (P_1) , that a strictly positive feasible solution is known, say $x^0 = e$, where $e \in R^n$ is a vector of ones, and that the optimal value of (P_1) is zero. Then it can be shown that (P_1) is equivalent to the homogeneous problem

$$\begin{aligned}
 & \min g^T x \\
 (HP_1) \quad & [A, -b] \begin{pmatrix} x \\ \xi \end{pmatrix} = 0 \\
 & x \geq 0, \xi \geq 0, (x, \xi) \neq 0.
 \end{aligned}$$

For any problem (P), let $F(P)$ denote the feasible region of (P) and $F_+(P)$ the set of strictly positive feasible solutions. We can evaluate any point $(x, \xi) \in F_+(HP_1)$ by Karmarkar's potential function

$$\begin{aligned}
 \phi_1(x, \xi) &= (n+1) \ln g^T x - \sum_j \ln x_j - \ln \xi \\
 &= \sum_j \ln \left(\frac{g^T x}{x_j} \right) + \ln \left(\frac{g^T x}{x_j} \right). \quad (2.1)
 \end{aligned}$$

Clearly, ϕ_1 is homogeneous of degree 0, and hence can be viewed as defined on positive rays, i.e., on certain points in projective space.

Karmarkar's algorithm generates a sequence $\{(x^k, \xi^k)\}$ in $F_+(HP_1)$ starting with $(x^0, \xi^0) = (e, 1)$, with

$$\phi_1(x^{k+1}, \xi^{k+1}) \leq \phi_1(x^k, \xi^k) - \epsilon \quad (2.2)$$

for some fixed positive ϵ and all k . Without loss of generality, we can take $\xi^k \equiv 1$, and then $\{x^k\}$ is a sequence in $F_+(P_1)$. We can then deduce from (2.2) that

$$g^T x^k \leq \left(\frac{e^T x^k}{n} \right) \exp \left(-\frac{k\epsilon}{n+1} \right) g^T x^0 \quad (2.3)$$

where e denotes the vector of ones in R^n . If $F(P_1)$ is bounded, we see that $g^T x^k$ converges linearly to the optimal value of zero.

If the data A , b and g are integer and the total input size is L , (2.3) can be used to show that $O(nL)$ iterations suffice to give an approximate solution from which an exact solution can be obtained by solving a system of equations. In general, $O(nq)$ iterations reduce the objective function by a factor of 2^q .

We now describe how the reduction in potential given by (2.2) is achieved. Let \hat{x} in $F_+(P_1)$ be given, and let $\Xi := \text{diag}(\hat{x})$ be the diagonal matrix whose diagonal entries are the components of \hat{x} . Let $\bar{g} := \Xi g$ and $\bar{A} := A\Xi$, and consider the rescaled problem

$$\begin{aligned}
 (\text{HP}_1) \quad & \min \bar{g}^T \bar{x} \\
 & [\bar{A}, -b] \begin{pmatrix} \bar{x} \\ \bar{\xi} \end{pmatrix} = 0 \\
 & \bar{x} \geq 0, \bar{\xi} \geq 0, (\bar{x}, \bar{\xi}) \neq 0.
 \end{aligned}$$

in terms of the rescaled variables $\bar{x} = \Xi^{-1}x$, $\bar{\xi} = \xi$. It is easy to see that, if (x, ξ) and $(\bar{x}, \bar{\xi})$ correspond as above, then

$$\begin{aligned}
 \phi_1(x, \xi) - \ell \ln \det \Xi &= \bar{\phi}_1(\bar{x}, \bar{\xi}) \\
 &:= (n+1)\ell \bar{g}^T \bar{x} - \sum_j \ell \ln \bar{x}_j - \ell \ln \bar{\xi},
 \end{aligned} \tag{2.4}$$

and that $(\hat{x}, 1)$ corresponds to the transformed vector $(e, 1)$. From (2.4) it is sufficient to obtain a constant reduction in $\bar{\phi}_1$ from the solution $(e, 1)$. We do this by making a step in the direction given by the negative of the projected gradient of $\bar{\phi}_1$. We find

$$\nabla \bar{\phi}_1 = \nabla \bar{\phi}_1(e, 1) = \begin{pmatrix} \frac{n+1}{\bar{g}^T e} \bar{g} - e \\ -1 \end{pmatrix}. \tag{2.5}$$

If P_M denotes projection into the null space of a matrix M , then our direction is

$$\begin{aligned}
 \bar{d} &:= -P_{[\bar{A}, -\bar{b}]} \nabla \bar{\phi}_1 \\
 &= -P_{[\bar{A}, -\bar{b}]} P_{[e^\tau, 1]} \begin{pmatrix} \frac{n+1}{\bar{g}^\tau e} \bar{g} \\ \bar{g}^\tau e \\ 0 \end{pmatrix} \\
 &= -\frac{n+1}{\bar{g}^\tau e} P_{\begin{bmatrix} \bar{A} & -\bar{b} \\ e^\tau & 1 \end{bmatrix}} \begin{pmatrix} \bar{g} \\ 0 \end{pmatrix} \tag{2.6}
 \end{aligned}$$

If we move from $(e, 1)$ to $(e, 1) + \alpha \bar{d} / \|\bar{d}\|$, we find

$$\begin{aligned}
 \bar{\phi}_1(e, 1) + \alpha \bar{d} / \|\bar{d}\| &= \bar{\phi}_1(e, 1) + \alpha \bar{d}^\tau \nabla \bar{\phi}_1 / \|\bar{d}\| + O(\alpha^2) \\
 &= \bar{\phi}_1(e, 1) - \alpha \|\bar{d}\| + O(\alpha^2). \tag{2.7}
 \end{aligned}$$

It is clear that a constant stepsize α can be chosen without violating positivity. Hence, since the higher order terms can be controlled, the reduction in $\bar{\phi}_1$ will be of the same order as the Euclidean norm of \bar{d} . In the proof of their lemma 3.1, Todd and Burrell [27] show that $\|\bar{d}\| \geq 1$. This yields a constant decrease in $\bar{\phi}_1$ per iteration, and hence convergence in $O(nL)$ iterations. Greater reduction in $\bar{\phi}_1$, and hence faster convergence, would follow if $\|\bar{d}\|$ could be shown to increase with n . We will return to this question in section 4.

After making a step in projective space to get a new $(\bar{x}, \bar{\xi})$, we normalize to $(\bar{x}/\bar{\xi}, 1)$ (note that $\bar{\phi}_1$ is unchanged), and then rescale to get our new feasible solution $\Xi \bar{x} / \bar{\xi}$ to (P_1) . Karmarkar's algorithm iterates this procedure.

An alternative method, the so-called affine variant (first proposed by Dikin [6] in 1967; see also [7], Barnes [3] and Vanderbei, Meketon and Freedman [29]), works directly in the original affine space. The rescaled variable $\bar{x} = \Xi^{-1}x$ is moved from e in the direction

$$\bar{d}_x = -P_{\bar{A}} \bar{g}. \quad (2.8)$$

Once again, we can use a constant stepsize α in the update $\bar{x} \leftarrow e + \alpha \bar{d}_x / \|\bar{d}_x\|$, and this will yield a decrease proportional to $\|\bar{d}_x\|$ in the objective function value. However, no lower bounds on $\|\bar{d}_x\|$ have been established and polynomial convergence has not been shown for this variant (and is believed unlikely; see Megiddo and Shub [19]).

Before we turn to path-following methods, we remark on a drawback of the potential function ϕ_1 . Assume that x is converging to the unique optimal solution x^* to (P_1) , and assume that this is a nondegenerate basic feasible solution. Let β ($\bar{\beta}$) index the basic (nonbasic) components of x^* . Let \hat{g} denote the reduced costs corresponding to the optimal basis, so that \hat{g}_j is positive for $j \notin \beta$. Then

$$\begin{aligned} \phi_1(x, 1) &= (n+1) \ln g^T x - \sum_j \ln x_j \\ &= (k+1) \ln g^T x + \sum_{j \in \bar{\beta}} \ln \left(\frac{\hat{g}_j^T x}{x_j} \right) - \sum_{j \in \beta} \ln x_j \end{aligned}$$

(there will be k basic indices); as $x \rightarrow x^*$, each term $\ln(\hat{g}_j^T x / x_j) \geq \ln \hat{g}_j$ ($j \in \bar{\beta}$) is bounded below, and each term $\ln x_j$ ($j \in \beta$) is converging to $\ln x_j^* > -\infty$. Hence a constant reduction in ϕ_1 implies an average reduction of a factor $(1 - \gamma/(k+1))$ for some constant γ in the objective function. However, this result is very dependent on primal and dual nondegeneracy. It is not clear that the factor $n+1$ multiplying $\ln g^T x$ gives the correct balance between reducing the objective function and staying away from the

constraint boundaries; and finding the right balance might necessitate knowing the degree of primal and dual degeneracy at the solution. We will see that this disadvantage vanishes for our method.

We now turn to an outline of path-following methods. We discuss a symmetric primal-dual variant, as in [21]. Associated with problem (P_1) is the barrier function problem with parameter μ :

$$\begin{aligned} \min \quad & g^T x - \mu \sum_j \ln x_j \\ (BP_1) \quad & Ax = b \\ & x > 0; \end{aligned}$$

under our assumptions, this will have a unique solution $x(\mu)$ for each $\mu > 0$, and $x(\mu)$ converges to an optimal solution of (P_1) as $\mu \rightarrow 0+$. The optimality conditions for (BP_1) , with Lagrange multipliers y , can be written as

$$\begin{aligned} g - \mu X^{-1} e - A^T y &= 0 \\ Ax &= b \end{aligned}$$

where X denotes $\text{diag}(x)$, and S below similarly denotes $\text{diag}(s)$. Letting $s := \mu X^{-1} e$, we can write these conditions as

$$XSe - \mu e = 0 \tag{2.9a}$$

$$Ax = b \tag{2.9b}$$

$$A^T y + s = g \tag{2.9c}$$

$$x > 0, \quad s > 0. \tag{2.9d}$$

Thus y , with associate slack vector s , is an interior solution to the dual problem

$$(D_1) \quad \begin{aligned} \max \quad & b^T y \\ & A^T y + s = g \\ & s \geq 0, \end{aligned}$$

and moreover the duality gap is $g^T x - b^T y = x^T s = e^T X S e = n\mu$. The equations (2.9) are also the optimality conditions for the barrier function problem

$$(BD_1) \quad \begin{aligned} \max \quad & b^T y + \mu \sum \ln s_j \\ & A^T y + s = g \\ & s > 0, \end{aligned}$$

with unique optimal solution $(y(\mu), s(\mu))$. Hence, solving (2.9) for a sequence of parameters μ approaching zero yields points on the curve $\{x(\mu)\}$ in the primal space and on the curve $\{(y(\mu), s(\mu))\}$ in the dual space. These are the central trajectories of Bayer and Lagarias [5] and the (central) pathways of Megiddo [18]. We call (x, y, s) or (x, s) centered if (2.9) holds for some μ .

Equation (2.9a) is nonlinear, so we must be content with approximate solutions. Assume we have a point $(\hat{x}, \hat{y}, \hat{s})$ satisfying (2.9b)–(2.9d) with

$$\|\hat{X}\hat{S}e - \hat{\zeta}e\| \leq \alpha\hat{\zeta} \quad (2.10)$$

where $\hat{\zeta} := e^T \hat{X}\hat{S}e/n$ and suitable $0 < \alpha < 1$; such a point we will call approximately centered. We then seek an approximate solution to (2.9) for some $\mu < \hat{\zeta}$ by taking a Newton step. The new point will be $(\hat{x} + d_x, \hat{y} + d_y, \hat{s} + d_s)$ where the direction satisfies the equations

$$\begin{aligned}
\hat{S}d_x + \hat{X}d_s &= -(\hat{X}\hat{S}e - \mu e) \\
Ad_x &= 0 \\
A^T d_y + d_s &= 0.
\end{aligned} \tag{2.11}$$

It can be shown (see, e.g., Monteiro and Adler [21]) that the result is also approximately centered, with duality gap approximately $n\mu$, if we choose $\mu = (1 - \gamma/\sqrt{n})\hat{\zeta}$ for some constant γ .

Most of the cited papers give little or no motivation or justification for the choice of μ above. However, Gonzaga [13] in his closely-related primal-only method, shows in section 3.3 that such a μ gives a constant bound on $\|\hat{X}^{-1}d_x\|$, thus ensuring that the new point is feasible. Kojima, Mizuno and Yoshise [15] set $\mu = \sigma\hat{\zeta}$ for some constant $\sigma < 1$ and find that they can only take a step size of order $1/n$, resulting in $O(nL)$ iterations. They also show (Theorem 2 and the following remark) that, even if one is exactly on the path $\{(x(\mu), y(\mu), s(\mu))\}$ and takes a step in the tangent direction for the maximum feasible distance, then the duality gap is reduced by at least the factor $(1 - 1/\sqrt{n})$.

3. SYMMETRIC PRIMAL AND DUAL PROBLEMS

In this section we reformulate (P_1) and (D_1) into symmetric forms, and define a symmetric combined primal-dual problem. Recall that we have

$$\begin{aligned}
(P_1) \quad & \min g^T x \\
& Ax = b \\
& x \geq 0,
\end{aligned}$$

satisfying (A1)–(A3). Since we assume (P_1) feasible, b is in the range of A , so that we can write

$$b = Ah \quad (3.1)$$

for suitable h . then the dual can be written

$$(D_1) \quad \begin{aligned} \max \quad & h^T A^T y \\ & A^T y + s = g \\ & s \geq 0. \end{aligned}$$

We prefer to write this in terms of s alone. Thus let B be a matrix whose rows span the null space of A ; by choosing a basis for this space we can assume that B is $\ell \times n$ with full row rank ℓ , with $k + \ell = n$. Then

$$A^T y + s = g \text{ for some } y \text{ iff } Bs = Bg$$

and moreover, the objective function of (D_1) can be written in terms of s alone since $h^T A^T y = g^T h - h^T s$. Thus we can rewrite (P_1) and (D_1) as

$$(P) \quad \begin{aligned} \min \quad & g^T x \\ & Ax = Ah \\ & x \geq 0 \end{aligned}$$

$$(D) \quad \begin{aligned} \min \quad & h^T s \\ & Bs = Bg \\ & s \geq 0 \end{aligned}$$

where the rows of A and B span complementary orthogonal subspaces of R^n , which we denote as

$$A \perp B. \quad (3.2)$$

The duality relation between (P) and (D) is easily deduced from the standard result: feasible solutions x and s have objective values satisfying

$$g^T x + h^T s \geq g^T h \quad (3.3)$$

and are optimal if and only if equality holds. Indeed, $x-h$ and $s-g$ lie in the null spaces of A and B respectively, and are therefore orthogonal, so

$$g^T x + h^T s - g^T h = x^T s, \quad (3.4)$$

which is nonnegative for feasible x, s .

The standard symmetric inequality-form linear programming problems are naturally included in our format above. If they are written as $\max \bar{c}^T \bar{x}$, $\bar{A} \bar{x} \leq \bar{b}$, $\bar{x} \geq 0$ and $\min \bar{b}^T \bar{y}$, $\bar{A}^T \bar{y} \geq \bar{c}$, $\bar{y} \geq 0$, then we set

$$A = (\bar{A}, I), \quad B = (I, -\bar{A}^T)$$

$$g = \begin{pmatrix} -\bar{c} \\ 0 \end{pmatrix} \quad \text{and} \quad h = \begin{pmatrix} 0 \\ \bar{b} \end{pmatrix} \quad (3.5)$$

to get corresponding instances (P) and (D).

We can combine (P) and (D) to get the primal-dual problem

$$\begin{aligned} \min \quad & g^T x + h^T s - g^T h \\ \text{(PD)} \quad & Ax \quad - Ah = 0 \\ & Bs - Bg = 0 \\ & x \geq 0, \quad s \geq 0 \end{aligned}$$

which has optimal value 0 when feasible. As stated, this problem is separable, but we may replace the objective function by $x^T s$ using (3.4). In this form, it is clear that (PD) is unaffected if we replace g by \tilde{g} and h by \tilde{h} where

$$A\tilde{h} = Ah \text{ and } B\tilde{g} = Bg. \quad (3.6)$$

Indeed, we have the useful

Lemma 1. Let (\tilde{P}) and (\tilde{D}) denote (P) and (D) defined using \tilde{g} and \tilde{h} in place of g and h , where (3.6) holds. Then (\tilde{P}) and (\tilde{D}) are equivalent to (P) and (D) in the sense that their feasible regions are unchanged and their objective functions the same up to additive constants.

Proof. It is clear that the feasible regions are unchanged. Now (3.6) implies that $\tilde{g} = g + A^T u$ and $\tilde{h} = h + B^T v$ for some u, v . Thus, for any feasible x ,

$$\tilde{g}^T x = g^T x + u^T A x = g^T x + u^T A h = g^T x + (\tilde{g}^T h - g^T h)$$

differs by a constant from $g^T x$. Similarly, $\tilde{h}^T s$ differs by a constant from $h^T s$ for feasible s .

We will use lemma 1 as follows; given $(\hat{x}, \hat{s}) \in F(PD)$, we can assume without loss of generality that $g = \hat{s}$ and $h = \hat{x}$.

Let us observe immediately one consequence of this. Suppose that $x = s = e$ is feasible in (PD). This solution is also centered, since (2.9a) holds with $\mu = 1$. Then we can take $g = h = e$ also. The direction chosen by the affine algorithm (see (2.8)) is then

$$d = \begin{pmatrix} d_x \\ d_s \end{pmatrix} = - \begin{pmatrix} P_A & 0 \\ 0 & P_B \end{pmatrix} \begin{pmatrix} e \\ e \end{pmatrix} = \begin{pmatrix} -P_A e \\ -P_B e \end{pmatrix}. \quad (3.7)$$

Recall that the difficulty with the affine variant was in bounding from below the norm of $\|\bar{d}_x\|$; even if $-\bar{g}$ were large (e.g., $\bar{g} = e$), its projection \bar{d}_x into the null space of \bar{A} could be small. Here this problem vanishes; since $A \perp B$, $d_x + d_s = -e$ and we can conclude that

$$\|d\| = \sqrt{n}. \quad (3.8)$$

Note that centering alone is not enough in this analysis—we need also the primal–dual formulation. In computing d from (3.7), the increased dimension implies no increase in work; given $d_x = -P_A e$, we find $d_s = -P_B e = -e - d_x$, since

$$P_A + P_B = I. \quad (3.9)$$

The importance of centering in the affine variant has been demonstrated also in Barnes and Jensen [4].

Since (PD) has optimal value 0, it is natural to consider its homogenization

$$\begin{aligned} \min \quad & g^T x + h^T s - g^T h \tau \\ \text{(HPD)} \quad & Ax - Ah\tau = 0 \\ & Bs - Bg\tau = 0 \\ & x \geq 0, \quad s \geq 0, \quad \tau \geq 0 \\ & (x, s, \tau) \neq 0. \end{aligned}$$

As for (HP_1) , we can view the feasible solutions of (HPD) as rays, i.e., as points in projective space. The relation equivalent to (3.4) for $(x, s, \tau) \in F(\text{HPD})$ with $\tau > 0$ is

$$g^T x + h^T s - g^T h \tau = x^T s / \tau. \quad (3.10)$$

To conclude this section, we observe that our assumptions:

- (A1) (P) has a strictly positive feasible solution;
- (A2) The set of optimal solutions of (P) is nonempty and bounded; and
- (A3) The matrix A has full row rank k ,

easily imply (B1) and (B2) below, and (B3) is without loss of generality:

- (B1) (D) has a strictly positive feasible solution;
- (B2) The set of optimal solutions of (D) is nonempty and bounded; and
- (B3) The matrix B has full row rank ℓ .

Hence (PD) has a strictly positive feasible solution, i.e. $F_+(PD) \neq \emptyset$, and also $F_+(HPD) \neq \emptyset$.

4. POTENTIAL FUNCTIONS AND SCALING

We associate with the homogeneous primal–dual problem the potential function

$$\phi(x, s, \tau) = \phi_\rho(x, s, \tau, g, h) \quad (4.1)$$

$$:= (n+\rho) \ell n(g^T x + h^T s - g^T h \tau) - \sum \ell n x_j - \sum \ell n s_j - (\rho-n) \ell n \tau$$

on $F_+(HPD)$. Note that this is homogeneous of degree 0, and hence is defined on (positive) rays. Perhaps the most natural choice of ρ is $n+1$, leading to the potential

function introduced by Karmarkar [14]. See also Ye and Todd [31]. However, we will find it profitable to consider other values for ρ , in particular $\rho \sim \sqrt{n}$. While it may seem strange to have a negative coefficient for $\ln \tau$ in (4.1), we note that (3.10) implies

$$\begin{aligned}\phi(x, s, \tau) &= (n + \rho) \ln x^T s - \sum \ln x_j - \sum \ln s_j - 2\rho \ln \tau \\ &= \rho \ln \left(\frac{x^T s}{\tau^2} \right) - \sum \ln \frac{x_j s_j}{x^T s}.\end{aligned}\quad (4.2)$$

The first equation shows that any $\rho > 0$ is reasonable, while the second gives a natural interpretation to ρ . Indeed, $x^T s / \tau^2$ is the objective function of the unscaled solution $(x/\tau, s/\tau)$ in $F_+(\text{PD})$, while $(x_j s_j / x^T s)$ is a positive vector lying on the unit simplex. Hence ρ balances a term that measures the objective function with the barrier function that seeks to keep $(x/\tau, s/\tau)$ centered in $F_+(\text{PD})$.

Our aim is to secure an appropriate decrease in ϕ at each iteration. To this end we will work with scaled problems.

Let Ξ be a positive definite diagonal $n \times n$ matrix and δ a positive scalar. Define the scaled data

$$\begin{aligned}\bar{A} &= \delta A \Xi, \quad \bar{B} = \delta B \Xi^{-1} \\ \bar{g} &= \delta \Xi g, \quad \bar{h} = \delta \Xi^{-1} h.\end{aligned}\quad (4.3)$$

Then, in terms of the rescaled variables

$$\bar{x} = \delta \Xi^{-1} x, \quad \bar{s} = \delta \Xi s, \quad \bar{\tau} = \tau \quad (4.4)$$

we have the equivalent problem

$$\begin{aligned}
 & \min \bar{g}^T \bar{x} + \bar{h}^T \bar{s} - \bar{g}^T \bar{h} \bar{\tau} \\
 (\text{HPD}) \quad & \bar{A} \bar{x} \quad \quad - \bar{A} \bar{h} \bar{\tau} = 0 \\
 & \bar{B} \bar{s} - \bar{b} \bar{g} \bar{\tau} = 0 \\
 & \bar{x} \geq 0, \bar{s} \geq 0, \bar{\tau} \geq 0 \\
 & (\bar{x}, \bar{s}, \bar{\tau}) \neq 0.
 \end{aligned}$$

We let $(\overline{\text{PD}})$ denote the corresponding inhomogeneous problem, obtained by replacing $\bar{\tau}$ by 1 in (HPD) . It follows from the definition (4.3) that

$$\bar{A} \perp \bar{B}; \quad (4.5)$$

also the objective function of $(\overline{\text{PD}})$ is δ^2 times that of (HPD) for corresponding points. If we set

$$\bar{\phi}(\bar{x}, \bar{s}, \bar{\tau}) := \phi_{\rho}(\bar{x}, \bar{s}, \bar{\tau}, \bar{g}, \bar{h}) \quad (4.6)$$

Then we find

$$\bar{\phi}(\bar{x}, \bar{s}, \bar{\tau}) = \phi(x, s, \tau) + \rho \ln \delta^2$$

whenever the arguments correspond as in (4.4). Thus to reduce the original potential function, it suffices to reduce $\bar{\phi}$ sufficiently for some appropriately scaled problem.

We now describe a particularly useful rescaling. Suppose we have $(\hat{x}, \hat{s}) \in F_+(PD)$, so that $(\hat{x}, \hat{s}, 1) \in F_+(HPD)$ with objective value $\hat{x}^\top \hat{s}$. Let

$$\hat{\zeta} := \hat{x}^\top \hat{s} / n. \quad (4.7)$$

According to lemma 1, we may take $g = \hat{s}$, $h = \hat{x}$ without changing the problems.

We also suppose (\hat{x}, \hat{s}) is approximately centered in $F_+(PD)$, so that

$$\|\hat{X}\hat{S}e - \hat{\zeta}e\| \leq \alpha \hat{\zeta} \quad (4.8)$$

for suitable $0 < \alpha < 1$. Now choose

$$\delta = \hat{\zeta}^{-1/2} \text{ and } \Xi = (\hat{X}\hat{S}^{-1})^{1/2}. \quad (4.9)$$

Then, with $g = \hat{s}$, $h = \hat{x}$, we find

$$\bar{g} = \bar{h} = \hat{\zeta}^{-1/2} (\hat{X}\hat{S})^{1/2} e =: \bar{e} \quad (4.10)$$

and

$$\bar{g}^\top \bar{h} = \hat{\zeta}^{-1} e^\top \hat{X}\hat{S}e = n. \quad (4.11)$$

Since \bar{e} is close to e , and our current point $(\hat{x}, \hat{s}, 1)$ has been rescaled to $(\bar{e}, \bar{e}, 1)$, we have simultaneously scaled the primal and dual problems reasonably well. At the same time, by employing a symmetric scaling for the primal and dual, we guarantee condition (4.5), which, as we shall see, is crucial.

5. PROJECTIONS AND THE SEARCH DIRECTION

We will be working with the scaled problem (HPD) and the associated potential function $\bar{\phi}$. The scaling will usually be that at the end of section 4, so that $\bar{g} = \bar{h}$, but we first compute the direction in general, assuming only that

$$\bar{g}^T \bar{h} = n \quad (5.1)$$

for notational simplicity; this can always be achieved by a scalar parameter. We also assume that \bar{g} and \bar{h} have been adjusted if necessary so that our current solution is

$$\begin{pmatrix} \bar{h} \\ \bar{g} \\ 1 \end{pmatrix}. \quad (5.2)$$

We first evaluate the gradient of the potential function at the current point. For a vector $u = (u_j) \in R^n$, it is convenient to denote

$$u^{-1} := (u_j^{-1}) \quad (5.3)$$

and to write u^{-T} for $(u^{-1})^T$. Then the gradient, using (4.2), is

$$\nabla \bar{\phi} = \nabla \bar{\phi}(\bar{h}, \bar{g}, 1) = \frac{n+\rho}{n} \begin{pmatrix} \bar{g} \\ \bar{h} \\ 0 \end{pmatrix} - \begin{pmatrix} \bar{h}^{-1} \\ \bar{g}^{-1} \\ 2\rho \end{pmatrix}. \quad (5.4)$$

The search direction we will employ is, as in projective methods for (P_1) , the projection of

$-\nabla\bar{\phi}$ into the tangent space of the feasible region. Hence, with

$$\bar{C} := \begin{bmatrix} \bar{A} & 0 & -\bar{A}\bar{h} \\ 0 & \bar{B} & -\bar{B}\bar{g} \end{bmatrix}, \quad (5.5)$$

we seek

$$\bar{d}' := -P_{\bar{C}} \nabla\bar{\phi}. \quad (5.6)$$

An important consequence of the primal–dual setting is that, with the scaling given at the end of section 4, \bar{d}' (or its modification \bar{d} below) cannot be too short.

Before computing \bar{d}' , we remark that, if the current point has $\bar{g} = \bar{h} = e$, $\tau = 1$, then the homogeneity of $\bar{\phi}$ implies $(e^\top, e^\top, 1)\nabla\bar{\phi} = 0$. Hence \bar{d}' is also the projection of $-\nabla\bar{\phi}$ into the null space of

$$\begin{bmatrix} \bar{A} & 0 & -\bar{A}\bar{h} \\ 0 & \bar{B} & -\bar{B}\bar{g} \\ e^\top & e^\top & 1 \end{bmatrix},$$

and this direction is also appropriate if the "simplex constraint" $e^\top\bar{x} + e^\top\bar{s} + \bar{\tau} = 2n+1$ is added to $(\overline{\text{HPD}})$. A scalar multiple of \bar{d}' is obtained if we project the negative gradient of the objective function into this same null space. However, if we work with $\nabla\bar{\phi}$, the extra appended row is unnecessary. Omitting the simplex constraint preserves the homogeneity of $(\overline{\text{HPD}})$ and has been propounded strongly by de Ghellinck and Vial [8,9]. In our context, adding a row of ones changes the direction since the current solution is generally not $\bar{x} = \bar{s} = e$, $\tau = 1$.

We now need to compute $P_{\tilde{C}}$. An important fact which follows from (4.5) is that

$$P_{\tilde{A}} + P_{\tilde{B}} = I; \quad (5.7)$$

we use this repeatedly. Let

$$T := \begin{pmatrix} \tilde{A} & 0 \\ 0 & \tilde{B} \end{pmatrix}, \quad u := \begin{pmatrix} -\tilde{h} \\ -\tilde{g} \end{pmatrix}, \quad \text{so } \tilde{C} = [T, Tu].$$

Then

$$P_T = \begin{pmatrix} P_{\tilde{A}} & 0 \\ 0 & P_{\tilde{B}} \end{pmatrix}, \quad I - P_T = \begin{pmatrix} P_{\tilde{B}} & 0 \\ 0 & P_{\tilde{A}} \end{pmatrix}, \quad \text{so}$$

$$v := (I - P_T) u = \begin{pmatrix} -P_{\tilde{B}} \tilde{h} \\ -P_{\tilde{A}} \tilde{g} \end{pmatrix}$$

Hence

$$\omega := (1 + v^T v)^{-1} = (1 + \tilde{h}^T P_{\tilde{B}} \tilde{h} + \tilde{g}^T P_{\tilde{A}} \tilde{g})^{-1}. \quad (5.8)$$

Now from section 4.2 of [26] we obtain

Lemma 2.

$$P_{\tilde{C}} = \begin{pmatrix} P_{\tilde{A}} & 0 & 0 \\ 0 & P_{\tilde{B}} & 0 \\ 0 & 0 & 0 \end{pmatrix} + \omega \begin{pmatrix} P_{\tilde{B}} \tilde{h} \\ P_{\tilde{A}} \tilde{g} \\ 1 \end{pmatrix} \begin{pmatrix} P_{\tilde{B}} \tilde{h} \\ P_{\tilde{A}} \tilde{g} \\ 1 \end{pmatrix}^T \quad (5.9)$$

where ω is given by (5.8).

To compute \bar{d}' in (5.6), we first calculate

$$\begin{pmatrix} P_{\bar{B}} \bar{h} \\ P_{\bar{A}} \bar{g} \\ 1 \end{pmatrix}^\top \begin{pmatrix} \bar{g} \\ \bar{h} \\ 0 \end{pmatrix} = \bar{g}^\top P_{\bar{B}} \bar{h} + \bar{g}^\top P_{\bar{A}} \bar{h} = \bar{g}^\top \bar{h} = n$$

using (5.7), and

$$\begin{pmatrix} P_{\bar{B}} \bar{h} \\ P_{\bar{A}} \bar{g} \\ 1 \end{pmatrix}^\top \begin{pmatrix} \bar{h}^{-1} \\ \bar{g}^{-1} \\ 2\rho \end{pmatrix} = \bar{h}^{-\top} P_{\bar{B}} \bar{h} + \bar{g}^{-\top} P_{\bar{A}} \bar{g} + 2\rho \\ = n + \rho + \sigma, \text{ where}$$

$$\sigma := \rho + \bar{g}^{-\top} P_{\bar{A}} \bar{g} + \bar{h}^{-\top} P_{\bar{B}} \bar{h} - n. \quad (5.10)$$

From (5.4), (5.6), (5.9) and the above, we find

$$\begin{aligned} \bar{d}' &= -P_{\bar{C}} \nabla \bar{\phi} = -\frac{n+\rho}{n} \left[\begin{pmatrix} P_{\bar{A}} \bar{g} \\ P_{\bar{B}} \bar{h} \\ 0 \end{pmatrix} + \omega n \begin{pmatrix} P_{\bar{B}} \bar{h} \\ P_{\bar{A}} \bar{g} \\ 1 \end{pmatrix} \right] \\ &\quad + \left[\begin{pmatrix} P_{\bar{A}} \bar{h}^{-1} \\ P_{\bar{B}} \bar{g}^{-1} \\ 0 \end{pmatrix} + \omega(n+\rho+\sigma) \begin{pmatrix} P_{\bar{B}} \bar{h} \\ P_{\bar{A}} \bar{g} \\ 1 \end{pmatrix} \right] \\ &= -\frac{n+\rho}{n} \begin{pmatrix} P_{\bar{A}} \bar{g} \\ P_{\bar{B}} \bar{h} \\ 0 \end{pmatrix} + \begin{pmatrix} P_{\bar{A}} \bar{h}^{-1} \\ P_{\bar{B}} \bar{g}^{-1} \\ 0 \end{pmatrix} + \omega \sigma \begin{pmatrix} P_{\bar{B}} \bar{h} \\ P_{\bar{A}} \bar{g} \\ 1 \end{pmatrix}. \end{aligned} \quad (5.11)$$

The direction \bar{d}' is appropriate to projective space, the domain of $(\overline{\text{HPD}})$. It is

convenient to transform this direction back to an appropriate direction for the inhomogeneous problem (PD). If we move from our current point a step size χ' in the direction $\bar{d}' = (\bar{d}'_x, \bar{d}'_s, \bar{d}'_\tau)$, we reach

$$\begin{pmatrix} \bar{h} + \chi' \bar{d}'_x \\ \bar{g} + \chi' \bar{d}'_s \\ 1 + \chi' \bar{d}'_\tau \end{pmatrix} = (1 + \chi' \bar{d}'_\tau) \begin{pmatrix} \bar{h} + \chi \bar{d}_x \\ \bar{g} + \chi \bar{d}_s \\ 1 \end{pmatrix},$$

where

$$\chi := \chi' / (1 + \chi' \bar{d}'_\tau) \text{ and}$$

$$\bar{d}_x = \bar{d}'_x - \bar{d}'_\tau \bar{h},$$

$$\bar{d}_s = \bar{d}'_s - \bar{d}'_\tau \bar{g}.$$

Given the homogeneity of $\bar{\phi}$, it is therefore equivalent to move a step size χ in the direction

$$\bar{d} = \begin{pmatrix} \bar{d}_x \\ \bar{d}_s \\ \bar{d}_\tau \end{pmatrix} := -\frac{n+\rho}{n} \begin{pmatrix} P_{\bar{A}} \bar{g} \\ P_{\bar{B}} \bar{h} \\ 0 \end{pmatrix} + \begin{pmatrix} P_{\bar{A}} \bar{h}^{-1} \\ P_{\bar{B}} \bar{g}^{-1} \\ 0 \end{pmatrix} + \omega \sigma \begin{pmatrix} -P_{\bar{A}} \bar{h} \\ -P_{\bar{B}} \bar{g} \\ 0 \end{pmatrix}. \quad (5.12)$$

We now note the simplifications that occur when, as in (4.10), $\bar{g} = \bar{h} = \bar{e}$. Then, from (5.1),

$$\tilde{e}^T \tilde{e} = n$$

and clearly

$$\tilde{e}^{-T} \tilde{e} = n.$$

Hence from (5.8)

$$\omega = (1 + \tilde{e}^T P_{\bar{A}} \tilde{e} + \tilde{e}^T P_{\bar{B}} \tilde{e})^{-1} = (1 + \tilde{e}^T \tilde{e})^{-1} = (n+1)^{-1}$$

and from (5.10)

$$\sigma = \rho + \tilde{e}^{-T} P_{\bar{A}} \tilde{e} + \tilde{e}^{-T} P_{\bar{B}} \tilde{e} - n = \rho + \tilde{e}^{-T} \tilde{e} - n = \rho.$$

Hence,

$$\bar{d} = -(1+\psi) \begin{pmatrix} P_{\bar{A}} \tilde{e} \\ P_{\bar{B}} \tilde{e} \\ 0 \end{pmatrix} + \begin{pmatrix} P_{\bar{A}} \tilde{e}^{-1} \\ P_{\bar{B}} \tilde{e}^{-1} \\ 0 \end{pmatrix} \quad (5.13)$$

where

$$\psi = \frac{2n+1}{n(n+1)} \rho. \quad (5.14)$$

Again, if \tilde{e} is close to e and a suitable value of ψ is chosen, our primal-dual setting assures us that \bar{d} cannot be too small. Further, there is no computational expense for this increased dimensionality; if

$$f := -(1+\psi)\tilde{e} + \tilde{e}^{-1}, \quad (5.15)$$

then

$$\bar{d}_x = P_{\bar{A}} f \text{ and } \bar{d}_s = f - \bar{d}_x. \quad (5.16)$$

To conclude this section, we relate our direction \bar{d} in (5.13) to those generated by path-following methods. Consider equations (2.11). If we scale as at the end of section 4, we can rewrite these equations in terms of

$$\bar{d}_x = \delta \Xi^{-1} d_x, \quad \bar{d}_s = \delta \Xi d_s.$$

The last equation of (2.11) is equivalent to $Bd_s = 0$. We thus obtain

$$(\hat{X}\hat{S})^{1/2}(\bar{d}_x + \bar{d}_s) = -\delta(\hat{X}\hat{S}e - \mu e)$$

$$\bar{A}\bar{d}_x = 0$$

$$\bar{B}\bar{d}_s = 0.$$

The first equation can be written

$$\bar{d}_x + \bar{d}_s = -\bar{e} + (\mu/\hat{\zeta})\bar{e}^{-1},$$

and so

$$\bar{d}_x = P_{\bar{A}} f', \quad \bar{d}_s = P_{\bar{B}} f', \quad (5.17)$$

with

$$f' = -\bar{e} + (\mu/\hat{\zeta})\bar{e}^{-1}. \quad (5.18)$$

The similarity to (5.15), (5.16) is striking. Indeed, if ψ is chosen so that $1 + \psi = \hat{\zeta}/\mu$, then the two directions coincide. If $\mu = \hat{\zeta}(1 - \gamma/\sqrt{n})$ for some constant γ , this suggests $\psi = \gamma'/\sqrt{n}$ for some constant γ' . Our choice of ψ will turn out to have this form, but it is motivated by obtaining suitable reductions in the potential function $\bar{\phi}$.

The analysis above is appropriate for the symmetric primal–dual algorithms of Monteiro and Adler [21] and Kojima, Mizuno and Yoshise [16]. Indeed, equations (2.11) are exactly equivalent to equations (3.1) of [21] with $\tilde{X} = \hat{X}$, $\tilde{Z} = \hat{S}$ and corresponding changes of notation. Kojima et al. [16] deal with the linear complementarity problem $w = Mz + q$ where M is positive semi-definite. Given the linear programming problems $\max \tilde{c}^T \tilde{x}$, $\tilde{A}\tilde{x} \leq \tilde{b}$, $\tilde{x} \geq 0$, $\min \tilde{b}^T \tilde{y}$, $\tilde{A}^T \tilde{y} \geq \tilde{c}$, $\tilde{y} \geq 0$, as above (3.5), we set

$$w = \begin{pmatrix} \tilde{u} \\ \tilde{v} \end{pmatrix}, \quad z = \begin{pmatrix} \tilde{y} \\ \tilde{x} \end{pmatrix}, \quad M = \begin{pmatrix} 0 & -\tilde{A} \\ \tilde{A}^T & 0 \end{pmatrix} \quad \text{and} \quad q = \begin{pmatrix} \tilde{b} \\ -\tilde{c} \end{pmatrix}$$

to get the corresponding linear complementarity problem, where \tilde{u} and \tilde{v} are primal and dual slack vectors. Then if (P) and (D) are defined using (3.5), with

$$x = \begin{pmatrix} \tilde{x} \\ \tilde{u} \end{pmatrix}, \quad s = \begin{pmatrix} \tilde{v} \\ \tilde{y} \end{pmatrix},$$

it is easy to see that Kojima et al.'s equations (2.1), which in the notation above are

$$Wd_z + Zd_w = ZWe - \mu e, \quad d_w = Md_z$$

can be rearranged to give our equations (2.11).

Finally, Gonzaga's direction \tilde{d}_N in his algorithm 4.2 [12] is of similar form if his scaling vector \tilde{z} is taken as $\zeta^{1/2}(\hat{X}\hat{S}^{-1})^{1/2}e \approx \hat{x}$. Indeed, in his notation

$$\tilde{d}_N = -\tilde{P}(\tilde{c}/\epsilon_{k+1} - Y^{-1}e),$$

and with the choice of \tilde{z} above, $\tilde{P} = P_{\tilde{A}}$ and $y = \tilde{Z}^{-1}\tilde{x} = \tilde{e}$, so $Y^{-1}e = \tilde{e}^{-1}$. Also,

$\tilde{c} = \tilde{Z}g$, and by the proof of lemma 1 its projection is the same as that of $\hat{\zeta}\tilde{e}$. Hence

$$\tilde{d}_N = P_{\tilde{A}}(-(\tilde{\zeta}/\epsilon_{k+1})\tilde{e} + \tilde{e}^{-1}),$$

which relates directly to (5.15)–(5.16).

6. THE ALGORITHM

We now have all the ingredients for our algorithm. We assume we have available a point $(x^0, s^0) \in F_+(PD)$ that is approximately centered, so that

$$\|X^0 S^0 e - \zeta^0 e\| \leq \alpha \zeta^0, \quad (6.1)$$

where $\zeta^0 = e^T X^0 S^0 e / n$, for suitable $0 < \alpha < 1$. Obtaining such a point has been discussed in [12,15,16,21–23,28]. We generate a sequence $\{(x^k, s^k)\} \subseteq F_+(PD)$ as follows.

Given (x^k, s^k) , set $\hat{x} = x^k$, $\hat{s} = s^k$ and define the scaled data by (4.3), with δ and Ξ given by (4.9). Then (x^k, s^k) is transformed into (\tilde{e}, \tilde{e}) . We choose \tilde{d} by (5.13) and set

$$\begin{pmatrix} \bar{x} \\ \bar{s} \end{pmatrix} := \begin{pmatrix} \tilde{e} \\ \tilde{e} \end{pmatrix} + \chi \begin{pmatrix} \tilde{d}_x \\ \tilde{d}_s \end{pmatrix} \quad (6.2)$$

for appropriate $\chi > 0$. Now we unscale the point to get

$$\begin{pmatrix} x^{k+1} \\ s^{k+1} \end{pmatrix} = \begin{pmatrix} \delta^{-1} \Xi \bar{x} \\ \delta^{-1} \Xi^{-1} \bar{s} \end{pmatrix}. \quad (6.3)$$

We aim to prove

Theorem 1. Let $\rho = \frac{2n+2}{2n+1} \sqrt{n}$ and $\psi = \frac{2}{\sqrt{n}}$. Let $\alpha = 1/3$ and $\chi = 1/15$. Then if (x^0, s^0) satisfies (6.1), the algorithm above is well-defined and generates a sequence of points satisfying

$$\|X^k S^k e - \zeta^k e\| \leq \alpha \zeta^k \quad (6.4)$$

where $\zeta^k = e^T X^k S^k e / n$ and

$$\phi(x^{k+1}, s^{k+1}, 1) \leq \phi(x^k, s^k, 1) - \beta \quad (6.5)$$

where $\beta = 1/9$.

The inequality (6.5) implies the desired complexity of the algorithm. Indeed, from (4.2)

$$\phi(x, s, 1) = \rho \ln(x^T s) - \sum_j \ln \frac{x_j s_j}{x^T s}$$

and (6.1) shows that the last term is bounded for (x^0, s^0) (and by (6.4) for all (x^k, s^k)). Hence, with $\rho \sim \sqrt{n}$, (6.5) gives

Corollary 1. In $O(\sqrt{n} L)$ steps, the algorithm above yields

$$e^T X^k S^k e \leq 2^{-L} e^T X^0 S^0 e. \quad (6.6)$$

In the integer model, a suitable choice for (x^0, s^0) will then show that $O(\sqrt{n} L)$ steps suffice to give a solution from which optimal solutions to (P) and (D) can be recovered. See [12,15,16,21–23,28]. Here L is the input size of the instance.

In fact, our analysis below will show (6.6) directly, by proving that the objective function decreases by a factor $(1 - \gamma/\sqrt{n})$ for constant γ at each iteration. However, since our algorithm is motivated by decreasing the potential function ϕ , we prefer to use it as a criterion. The approximate centering shown by (6.4) is an automatic byproduct of the method.

Towards proving theorem 1, we note that conditions (6.4)–(6.5) are invariant under scaling of the problem, so that it is sufficient to show that the move from (\tilde{e}, \tilde{e}) to (\bar{x}, \bar{s}) preserves (6.4) and yields the potential function reduction (6.5). Removing the overbars for notational simplicity, we only need to establish the following (here \tilde{e}^2 denotes the vector with $(\tilde{e}^2)_j = (\tilde{e}_j)^2$):

Lemma 3. Let ρ, ψ, α, χ and β be as in the theorem. Suppose

$$\|\tilde{e}^2 - e\| \leq \alpha. \quad (6.7)$$

If

$$d = \begin{pmatrix} d_x \\ d_s \\ d_r \end{pmatrix} = -(1+\psi) \begin{pmatrix} P_A \tilde{e} \\ P_B \tilde{e} \\ 0 \end{pmatrix} + \begin{pmatrix} P_A \tilde{e}^{-1} \\ P_B \tilde{e}^{-1} \\ 0 \end{pmatrix} \quad (6.8)$$

and

$$\begin{pmatrix} x \\ s \end{pmatrix} := \begin{pmatrix} \tilde{e} \\ \tilde{e} \end{pmatrix} + \chi \begin{pmatrix} d_x \\ d_s \end{pmatrix},$$

then

$$\|XSe - \zeta e\| \leq \alpha\zeta \quad (6.9)$$

with $\zeta = \mathbf{x}^\top \mathbf{s}/n$, and

$$\phi(\mathbf{x}, \mathbf{s}, 1) \leq \phi(\bar{\mathbf{e}}, \bar{\mathbf{e}}, 1) - \beta. \quad (6.10)$$

7. ANALYSIS

This section proves lemma 3. We carry out the analysis as far as possible using general values for ρ , ψ , α , χ and β . This will help us discuss the choice of ρ , ψ made. Only when we complete the proof of lemma 3 will the specific values be used. However, we assume throughout that χ is such that $(\mathbf{x}, \mathbf{s}) > 0$. Since $(\bar{\mathbf{e}}, \bar{\mathbf{e}})$ is feasible and $\mathbf{d}_\mathbf{x}$ ($\mathbf{d}_\mathbf{s}$) in the null space of \mathbf{A} (\mathbf{B}), this assures us that (\mathbf{x}, \mathbf{s}) is feasible also.

We now calculate some key quantities. As usual, we denote by $\mathbf{D}_\mathbf{x}$ and $\mathbf{D}_\mathbf{s}$ $\text{diag}(\mathbf{d}_\mathbf{x})$ and $\text{diag}(\mathbf{d}_\mathbf{s})$ respectively, and $\tilde{\mathbf{E}}$ denotes $\text{diag}(\bar{\mathbf{e}})$.

Lemma 4. We have

$$-\nabla \phi^\top \mathbf{d} = \rho \psi + \bar{\mathbf{e}}^{-\top} \bar{\mathbf{e}}^{-1} - n, \quad (7.1)$$

$$\mathbf{d}^\top \mathbf{d} = n\psi^2 + \bar{\mathbf{e}}^{-\top} \bar{\mathbf{e}}^{-1} - n, \quad (7.2)$$

$$\zeta := \mathbf{x}^\top \mathbf{s}/n = 1 - \chi\psi, \text{ and} \quad (7.3)$$

$$\mathbf{X}\mathbf{S}\mathbf{e} - \zeta \mathbf{e} = (1 - \chi(1+\psi))(\bar{\mathbf{e}}^2 - \mathbf{e}) + \chi^2 \mathbf{D}_\mathbf{x} \mathbf{D}_\mathbf{s} \mathbf{e}. \quad (7.4)$$

Proof. In the present setting, (5.4) simplifies to

$$\nabla \phi = \frac{n+\rho}{n} \begin{pmatrix} \bar{\mathbf{e}} \\ \bar{\mathbf{e}} \\ 0 \end{pmatrix} - \begin{pmatrix} \bar{\mathbf{e}}^{-1} \\ \bar{\mathbf{e}}^{-1} \\ 2\rho \end{pmatrix}.$$

Then (7.1) and (7.2) are straightforward consequences of (6.8), using $P_A + P_B = I$ repeatedly.

Next note that $d_x^T d_s = 0$, since d_x and d_s lie in the null spaces of A and B respectively, while

$$d_x + d_s = -(1+\psi)\tilde{e} + \tilde{e}^{-1}. \quad (7.5)$$

Hence

$$\begin{aligned} \zeta &= (\tilde{e} + \chi d_x)^T (\tilde{e} + \chi d_s) / n \\ &= (\tilde{e}^T \tilde{e} + \chi \tilde{e}^T (d_x + d_s)) / n \\ &= (n + \chi(-n(1+\psi) + n)) / n \\ &= 1 - \chi\psi, \end{aligned}$$

giving (7.3), and

$$\begin{aligned} XSe - \zeta e &= (\tilde{E} + \chi D_x)(\tilde{E} + \chi D_s)e - \zeta e \\ &= \tilde{e}^2 + \chi \tilde{E}(d_x + d_s) + \chi^2 D_x D_s e - (1 - \chi\psi)e \\ &= \tilde{e}^2 + \chi(-(1+\psi)\tilde{e}^2 + e) + \chi^2 D_x D_s e - (1 - \chi\psi)e \\ &= (1 - \chi(1+\psi))(\tilde{e}^2 - e) + \chi^2 D_x D_s e, \end{aligned}$$

establishing (7.4).

To use Lemma 4, we need to bound $\tilde{e}^{-T} \tilde{e}^{-1}$.

Lemma 5. If

$$\|\tilde{e}^2 - e\| \leq \alpha$$

with $\alpha \leq 1/3$, then

$$n \leq \tilde{e}^{-\tau} \tilde{e}^{-1} \leq n+1. \quad (7.6)$$

Proof. For the lower bound, note that $\tilde{e}^{-\tau} \tilde{e} = n$, while $\|\tilde{e}\| = \sqrt{n}$. Hence $\|\tilde{e}^{-1}\| \geq \sqrt{n}$.

Now since $\tilde{e} > 0$, we certainly have

$$\|\tilde{e} - e\| \leq \alpha$$

also. So $\|\tilde{e}^{-1} - e\| = \|\tilde{E}^{-1}(e - \tilde{e})\| \leq \frac{1}{1-\alpha} \|e - \tilde{e}\| \leq \frac{\alpha}{1-\alpha}$, since each diagonal entry of \tilde{E}^{-1} is at most $(1-\alpha)^{-1}$. Hence $\|\tilde{e}^{-1} - \tilde{e}\| \leq \alpha + \frac{\alpha}{1-\alpha} < 1$ since $\alpha \leq 1/3$. Finally $\|\tilde{e}^{-1}\|^2 = \|\tilde{e} + (\tilde{e}^{-1} - e)\|^2 = \|\tilde{e}\|^2 + \|\tilde{e}^{-1} - \tilde{e}\|^2 \leq n+1$ since $\tilde{e}^{\tau}(\tilde{e}^{-1} - \tilde{e}) = n - n = 0$.

Finally, we need

Lemma 6. If $1 - \alpha - \chi\|d\| > 0$, then

$$\phi(x, s, 1) \leq \phi(\tilde{e}, \tilde{e}, 1) + \chi \nabla \phi^{\tau} d + \frac{1}{2(1-\alpha-\chi\|d\|)} \chi^2 d^{\tau} d. \quad (7.7)$$

Proof. We start with

$$\begin{aligned} \ln(1+\lambda) &= \lambda - \frac{\lambda^2}{2} + \frac{\lambda^3}{3} \dots \\ &\geq \lambda - \sum_{i=2}^{\infty} \frac{|\lambda|^i}{i} = \lambda - \frac{1}{2(1-|\lambda|)} \lambda^2 \end{aligned}$$

if $|\lambda| < 1$. Now suppose $|\epsilon^2 - 1| \leq \alpha$ (so $|\epsilon - 1| \leq \alpha$ also) and $|\theta| \leq \|d\|$. Then

$$\begin{aligned}\ln(\epsilon + \chi\theta) &= \ln\epsilon + \ln(1 + \chi\frac{\theta}{\epsilon}) \\ &\geq \ln\epsilon + \chi\frac{\theta}{\epsilon} - \frac{1}{2(1 - \chi|\frac{\theta}{\epsilon}|)\epsilon^2} \chi^2\theta^2\end{aligned}$$

Now if $\epsilon \leq 1$, then

$$\begin{aligned}(1 - \chi|\frac{\theta}{\epsilon}|)\epsilon^2 &= \epsilon^2 - \chi|\theta|\epsilon \geq \epsilon^2 - \chi|\theta| \\ &\geq 1 - \alpha - \chi\|d\|,\end{aligned}$$

while if $\epsilon > 1$, we have

$$\begin{aligned}(1 - \chi|\frac{\theta}{\epsilon}|)\epsilon^2 &\geq 1 - \chi|\theta|\epsilon \geq 1 - \chi|\theta|(1 + \alpha) \\ &\geq 1 - \alpha - \chi\|d\|\end{aligned}$$

since $\chi|\theta| < 1$. Hence

$$\ln(\epsilon + \chi\theta) \geq \ln\epsilon + \chi\frac{\theta}{\epsilon} - \frac{1}{2(1 - \alpha - \chi\|d\|)} \chi^2\theta^2. \quad (7.8)$$

Note that the first two terms are the first-order Taylor approximation of $\ln(\epsilon + \chi\theta)$.

Now consider

$$\phi(x, s, 1) = \rho \ln(\bar{e}^T x + \bar{e}^T s - n) - \sum \ln x_j - \sum \ln s_j.$$

The first term is concave, hence its first-order Taylor approximation is an overestimate.

For the remaining terms we use the approximation given by (7.8), where ϵ denotes a component of \tilde{e} and θ a component of d_x or d_s . Since the sum of all such θ^2 is $d^\top d$, the result follows.

Proof of Lemma 3. By lemmas 4 and 5, $4 \leq d^\top d \leq 5$ and so $\|d\| \leq 5/2$ and $1 - \alpha - \chi\|d\| \geq 1 - \frac{1}{3} - \frac{1}{15} \cdot \frac{5}{2} = \frac{1}{2}$. Thus each component of x and of s is positive. Also, $\chi(1 + \psi) < 1$, so that, again by lemma 4

$$\begin{aligned} \|XSe - \zeta e\| &\leq (1 - \chi(1 + \psi))\|\tilde{e}^2 - e\| + \chi^2 d^\top d \\ &\leq (1 - \chi\psi)\alpha + \chi(\chi d^\top d - \alpha) \\ &\leq \alpha\zeta. \end{aligned} \tag{7.9}$$

Finally, by lemmas 4 and 6 and the above estimates,

$$\begin{aligned} \phi(x, s, 1) &\leq \phi(\tilde{e}, \tilde{e}, 1) + \chi \nabla \phi^\top d + \frac{1}{2(1 - \alpha - \chi\|d\|)} \chi^2 d^\top d \\ &\leq \phi(\tilde{e}, \tilde{e}, 1) - \frac{1}{15} \cdot 2 + \frac{1}{2 \cdot (\frac{1}{2})} \left(\frac{1}{15}\right)^2 \cdot 5 \\ &= \phi(\tilde{e}, \tilde{e}, 1) - \frac{1}{9}. \end{aligned}$$

This completes the proof.

8. DISCUSSION

We now discuss the choice of parameters made. From (5.14), ψ is about $2\rho/n$, so to achieve a good reduction in the potential function we would like $-\nabla \phi^\top d$ and hence ρ large. To maintain feasibility, we need $\chi = O(1/\|d\|)$. Let us examine the consequences

of choosing ρ of order n . Then ψ is of order 1, $-\nabla\phi^T d$ of order n , and $\|d\|$ of order \sqrt{n} . Hence χ is of order $1/\sqrt{n}$, and the first order change in ϕ of order \sqrt{n} . Moreover, (7.7) shows that the error in this first-order estimate is of order 1. Hence we can achieve a decrease in ϕ of order \sqrt{n} (see also Ye and Todd [31]). This would seem to imply a complexity of $O(\sqrt{n} L)$, since with ρ of order n we need a reduction of order n in ϕ to reduce the objective function by a constant; this bound also appears to follow from the objective function reduction in (7.3). However, such a choice of ρ does not seem to allow the algorithm to remain approximately centered, and hence the improvement cannot be sustained. While the first-order term in (7.4) is very attractive, the second-order term seems to be too large. Indeed, (7.9) shows that we require $\chi d^T d - \alpha < 0$, and so χ should be $O(1/\|d\|^2)$. This would give χ of order $1/n$ and hence only a constant reduction in ϕ and a reduction in the objective function of $(1-\gamma/n)$ for constant γ , as in Karmarkar's algorithm or Kojima et al. [15].

On the other hand, if ρ is smaller than order \sqrt{n} , then $-\nabla\phi^T d$ is smaller than order 1, and we may not be able to achieve a reduction in ϕ due to second order terms. Our choice of ρ as approximately \sqrt{n} balances these requirements nicely.

We must mention a disadvantage of our choice of parameters. One of the goals of our research was to develop an algorithm that required only $\sqrt{n} L$ iterations as in the path-following methods and yet gave the possibility that line searches could reduce the complexity further. While line searches are possible in the path-following algorithms [12], the direction is determined directly or indirectly by seeking a reduction of the objective function by a factor $(1 - \gamma/\sqrt{n})$ for constant γ , and it is not clear that these directions are good for longer steps. Our directions arise from seeking reductions in a potential function, and one could hope that line searches would be more effective.

Unfortunately, while feasibility demands only χ to be $O(1/\|d\|_\infty)$, which might be of order \sqrt{n} , the approximate centering places more stringent conditions. Indeed, (7.4)

suggests that we must maintain $1 - \chi(1+\psi)$ positive (or not too negative) which limits χ to order 1. It may well be that the inherent curvature of the "cone" of approximate centers does not allow line searches to be effective. Then either one must establish a method that successfully moves further down the path of centers, or employ a higher-order predictor of the path, as suggested by Bayer and Lagarias [5] and Adler, Karmarkar, Resende and Veiga [1]. These remarks suggest that the bi-directional search proposed by Tanabe [25] and Gonzaga [13], which amounts to varying ψ in the definition of the direction \bar{d} of (5.13), may not be as successful as hoped.

Finally, while we hope that our framework has shown the advantages of primal-dual setting and approximate centering in projective algorithms, the true worst-case complexity of Karmarkar's primal-only projective algorithm is still unknown. Are there examples requiring $O(nL)$ iterations, or can the improved behavior observed in practice be proved to hold in general? An interesting discussion relating to this question can be found in McDiarmid [17].

ACKNOWLEDGEMENT

This research benefited from several very helpful conversations with Kurt Anstreicher.

REFERENCES

- [1] I. Adler, N. Karmarkar, M. Resende and G. Veiga, "An implementation of Karmarkar's algorithm for linear programming," Report ORC86-8, Operations Research Center, University of California (Berkeley, CA, 1986).
- [2] K. Anstreicher, "The worst-case step in Karmarkar's algorithm," manuscript, Yale School of Organization and Management (New Haven, CT, 1986).
- [3] E.R. Barnes, "A variation on Karmarkar's algorithm for solving linear programming problems," Mathematical Programming 36 (1986) 174-182.
- [4] E.R. Barnes and D.L. Jensen, "A polynomial-time version of the affine-scaling algorithm," manuscript in preparation, IBM T.J. Watson Research Center (Yorktown Heights, NY, 1987).
- [5] D. Bayer and J.C. Lagarias, "The nonlinear geometry of linear programming I. Affine and projective scaling trajectories, II. Legendre transform coordinates and central trajectories," manuscripts, to appear in Transactions of the American Mathematical Society (1987).
- [6] I.I. Dikin, "Iterative solution of problems of linear and quadratic programming," Soviet Mathematics Doklady 8 (1967), 674-675.
- [7] I.I. Dikin, "On the speed of an iterative process," Upravlyaemye Sistemi 12 (1974), 54-60 (in Russian).
- [8] G. de Ghellinck and J.-P. Vial, "An extension of Karmarkar's algorithm for solving a system of linear homogeneous equations on the simplex," CORE discussion paper 8538, (Louvain, Belgium, 1985, revised 1986). To appear in Mathematical Programming.
- [9] G. de Ghellinck and J.-P. Vial, "A polynomial Newton method for linear programming," Algorithmica 1 (1986) 425-453.
- [10] A.V. Fiacco and G.P. McCormick. Nonlinear Programming: Sequential Unconstrained Minimization Techniques (John Wiley and Sons, New York, 1968).
- [11] K.R. Frisch, "The Logarithmic Potential Method of Convex Programming," unpublished manuscript, University Institute of Economics (Oslo, Norway, 1955).
- [12] C.C. Gonzaga, "An algorithm for solving linear programming problems in $O(n^3L)$ operations," Memorandum Number UCB/ERL M87/10, Electronics Research Laboratory, University of California (Berkeley, CA, 1987).
- [13] C.C. Gonzaga, "Search directions for interior linear programming methods," memorandum No. UCB/ERL M87/44, Electronics Research Laboratory, University of California (Berkeley, CA, 1987).
- [14] N. Karmarkar, "A new polynomial time algorithm for linear programming," Combinatorica 4 (1984) 373-395.

- [15] M. Kojima, S. Mizuno and A. Yoshise, "A primal-dual interior point method for linear programming," Research report no. B-188, Department of Information Sciences, Tokyo Institute of Technology (Tokyo, Japan, 1987).
- [16] M. Kojima, S. Mizuno and A. Yoshise, "A polynomial-time algorithm for a class of linear complementarity problems," Research report no. B-193, Department of Information Sciences, Tokyo Institute of Technology (Tokyo, Japan, 1987).
- [17] C.J.H. McDiarmid, "On the improvement per iteration in Karmarkar's algorithm for linear programming," manuscript, Institute of Economics and Statistics, Oxford University (Oxford, U.K., 1986).
- [18] N. Megiddo, "Pathways to the optimal set in linear programming," Research report RJ 5295, IBM Almaden Research Center (San Jose, CA, 1986).
- [19] N. Megiddo and M. Shub, "Boundary behaviour of interior point algorithms in linear programming," Research report RJ 5319, IBM Thomas J. Watson Research Center (Yorktown Heights, NY, 1986).
- [20] J.E. Mitchell and M.J. Todd, "On the relationship between the search directions in the affine and projective variants of Karmarkar's linear programming algorithm," Technical Report no. 725, School of Operations Research, Cornell University (Ithaca, NY, 1986).
- [21] R.C. Monteiro and I. Adler, "An $O(n^3L)$ primal-dual interior point algorithm for linear programming," manuscript, Department of Industrial Engineering and Operations Research, University of California (Berkeley, CA, 1987).
- [22] R.C. Monteiro and I. Adler, "An $O(n^3L)$ interior point algorithm for convex quadratic programming," manuscript, Department of Industrial Engineering and Operations Research, University of California (Berkeley, CA, 1987).
- [23] J. Renegar, "A polynomial-time algorithm based on Newton's method for linear programming," Report MSRI 07 118-86, Mathematical Sciences Research Institute (Berkeley, CA, 1986). To appear in Mathematical Programming 40 (1988).
- [24] G. Sonnevend, "An 'Analytic Center' for polyhedrons and new classes of global algorithms for linear (smooth, convex) programming," Proceedings of the 12th IFIP Conference on System Modeling and Optimization, Budapest, 1985, to appear in Lecture Notes in Control and Information Sciences (Springer-Verlag).
- [25] K. Tanabe, "Complementarity-enforcing centered Newton method for mathematical programming," manuscript, The Institute of Statistical Mathematics (Minatoku, Tokyo, 1987).
- [26] M.J. Todd, "Exploiting special structure in Karmarkar's linear programming algorithm," Technical Report 721, School of Operations Research and Industrial Engineering, Cornell University (Ithaca, NY, 1986). To appear in Mathematical Programming 41 (1988).
- [27] M.J. Todd and B.P. Burrell, "An extension of Karmarkar's algorithm for linear programming using dual variables," Algorithmica 1 (1986) 409-424.

- [28] P.M. Vaidya, "An algorithm for linear programming which requires $O(((m+n)n^2 + (m+n)^{1.5})L)$ arithmetic operations," preprint, AT&T Bell Laboratories (Murray Hill, NJ, 1987).
- [29] R.J. Vanderbei, M.S. Meketon and B.A. Freedman, "A modification of Karmarkar's linear programming algorithm," Algorithmica 1 (1986) 395–407.
- [30] Y. Ye, "Interior algorithms for linear, quadratic, and linearly constrained convex programming," Ph.D. dissertation, Department of Engineering–Economic Systems, Stanford University (Stanford, CA, 1987).
- [31] Y. Ye and M.J. Todd, "Containing and shrinking ellipsoids in the path-following algorithm," manuscript, Department of Engineering–Economic Systems, Stanford University (Stanford, CA, 1987).